

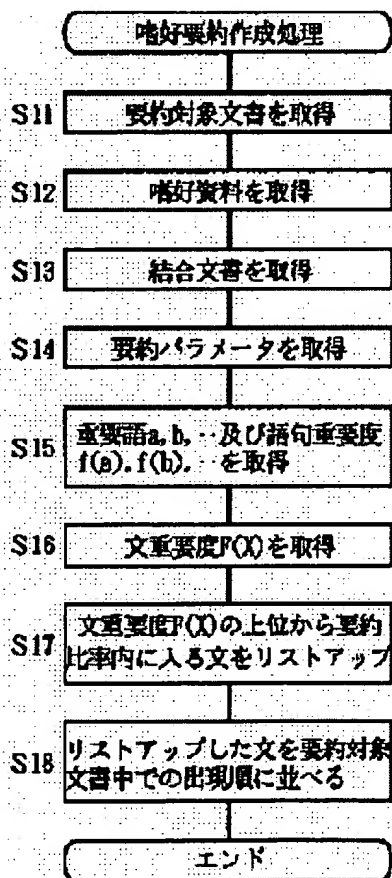
**DOCUMENT PROCESSOR, STORAGE MEDIUM STORING DOCUMENT
PROCESSING PROGRAM AND DOCUMENT PROCESSING METHOD**

Patent number: JP11045290
 Publication date: 1999-02-16
 Inventor: NOMURA NAOYUKI
 Applicant: JUST SYST CORP
 Classification:
 - International: G06F17/30; G06F17/27
 - european:
 Application number: JP19970218231 19970728
 Priority number(s): JP19970218231 19970728

Report a data error here

Abstract of JP11045290

PROBLEM TO BE SOLVED: To provide a document processor capable of preparing a summary based on the preference of a user such as a utilization purpose or the like, a storage medium storing a document processing program and a document processing method. **SOLUTION:** A document reflecting the preference of the user is combined to a summary object document and the candidate words of an important word are extracted from the entire obtained combined document by morpheme analysis or the like. Then, from an appearing frequency or the like in the combined document, the phrase importance $f(x)$ of the candidate word (x) is obtained and the candidate word of high phrase importance is turned to the important word. The obtained important words (a), (b),... and the phrase importance $f(a)$, $f(b)$... reflect the preference of the user more than the case of obtaining the important word and the phrase importance only from the summary object document. Then, based on the important words (a), (b),... and the phrase importance $f(a)$, $f(b)$..., the sentence importance $F(X)$ of the respective sentences of the summary object document is obtained and the sentences of the high sentence importance $F(X)$ are listed up, arranged in an appearing order in the summary object document and turned to the summary.



Data supplied from the esp@cenet database - Worldwide

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平11-45290

(43) 公開日 平成11年(1999) 2月16日

(51) Int.Cl.⁶

識別記号

F I

G 0 6 F 17/30

G 0 6 F 15/401

3 2 0 A

17/27

15/20

5 5 0 A

15/40

3 7 0 A

15/403

3 4 0 A

審査請求 未請求 請求項の数 7 F D (全 8 頁)

(21) 出願番号 特願平9-218231

(71) 出願人 390024350

株式会社ジャストシステム

徳島県徳島市沖浜東 3-46

(22) 出願日 平成9年(1997) 7月28日

(72) 発明者 野村 直之

徳島県徳島市沖浜東 3丁目46番地 株式会

社ジャストシステム内

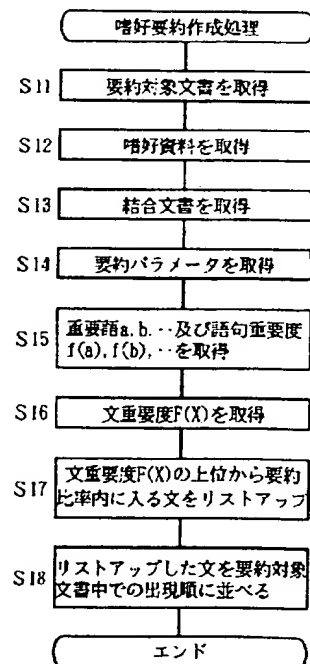
(74) 代理人 弁理士 川井 隆 (外 1 名)

(54) 【発明の名称】 文書処理装置、文書処理プログラムが記憶された記憶媒体、及び文書処理方法

(57) 【要約】

【課題】 本発明は、利用目的等のユーザーの嗜好を踏まえた要約の作成が可能な、文書処理装置、文書処理プログラムが記憶された記憶媒体、及び文書処理方法を提供すること。

【解決手段】 要約対象文書に、ユーザーの嗜好を反映した文書を結合し、得られた結合文書全体から形態素解析等により重要語の候補語を抽出する。そして、結合文書における出現頻度等から、候補語 x の語句重要度 $f(x)$ を取得し、語句重要度の高い候補語を重要語とする。得られた重要語 a, b, ... とその語句重要度 $f(a)$, $f(b)$, ... は要約対象文書のみから重要語や語句重要度を取得する場合よりも、ユーザーの嗜好の反映されたものとなる。そしてこの重要語 a, b, ... 及び語句重要度 $f(a)$, $f(b)$, ... に基づいて、要約対象文書の各文の文重要度 $F(X)$ を取得し、文重要度 $F(X)$ の高い文をリストアップし、要約対象文書中の出現順に並べて、要約とする。



【特許請求の範囲】

【請求項1】 要約の作成対象となる要約対象文書を取得する対象文書取得手段と、

ユーザーの嗜好を反映した嗜好資料を取得する嗜好資料取得手段と、

上記対象文書取得手段により取得した要約対象文書と上記嗜好資料取得手段により取得した嗜好資料とを結合して結合文書を取得する文書結合手段と、

上記文書結合手段により取得した結合文書から重要語句を抽出する重要語句抽出手段と、

前記重要語句抽出手段により取得された重要語句を用いて前記要約対象文書から重要文を選択する重要文選択手段と、

前記重要文選択手段により選択された重要文により前記文書の要約を作成する嗜好要約作成手段とを具備することを特徴とする文書処理装置。

【請求項2】 前記嗜好資料取得手段は、嗜好資料として、ユーザーの嗜好を反映した文書またはプロフィールを使用することを特徴とする請求項1に記載の文書処理装置。

【請求項3】 要約対象文書全体に対する要約の比率を取得する要約比率取得手段を備え、

前記重要文選択手段は、前記要約比率取得手段で取得した前記比率に従って重要文を選択することを特徴とする請求項1または請求項2に記載の文書処理装置。

【請求項4】 要約の作成対象となる要約対象文書を取得する対象文書取得機能と、

ユーザーの嗜好を反映した嗜好資料を取得する嗜好資料取得機能と、

上記対象文書取得機能により取得した要約対象文書と上記嗜好資料取得機能により取得した嗜好資料とを結合して結合文書を取得する文書結合機能と、

上記文書結合機能により取得した結合文書から重要語句を取得する重要語句取得機能と、

前記重要語句取得機能により取得された重要語句を用いて前記要約対象文書から重要文を選択する重要文選択機能と、

前記重要文選択機能により選択された重要文により前記要約対象文書の要約を作成する嗜好要約作成機能と、をコンピュータに実現させるためのコンピュータ読みとり可能な文書処理プログラムが記憶された記憶媒体。

【請求項5】 前記嗜好資料取得機能は、嗜好資料として、ユーザーの嗜好を反映した文書またはプロフィールを使用することを特徴とする請求項4に記載の文書処理プログラムが記憶された記憶媒体。

【請求項6】 要約対象文書全体に対する要約の比率を取得する要約比率取得機能を備え、

前記重要文選択機能は、前記要約比率取得機能で取得した前記比率に従って重要文を選択することを特徴とする請求項4または請求項5に記載の文書処理プログラムが

記憶された記憶媒体。

【請求項7】 要約の作成対象となる要約対象文書及びユーザーの嗜好を反映した嗜好資料を取得し、

上記要約対象文書と上記嗜好資料取得手段により取得した嗜好資料とを結合して結合文書を取得し、

上記結合文書から重要語句を取得し、

前記重要語句を用いて前記要約対象文書から重要文を選択し、

この重要文により前記文書の要約を作成することを特徴とする文書処理方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、文書処理装置、文書処理プログラムが記憶された記憶媒体、及び文書処理方法に関し、更に詳細には、利用目的等のユーザーの嗜好を踏まえた要約の作成に関する。

【0002】

【従来の技術】従来、書籍、論文、報告書等の各種の文書に対し、要約(抄録を含む)の自動作成処理をコンピュータを用いて行うことが行われている。文書の自動要約については、例えば、「全文情報からの意味的情報の抽出と加工」(情報処理学会第38回全国大会予稿集、第222頁;1989年)で提案されている。この方法では、まず文書中の重要語句を字種や動詞等の情報から抽出し、さらに重要語句の出現頻度から最重要語句を取得する。次に重要語句と最重要語句が出現するか否かから重要文を取得することで、自動的に要約を作成することが可能になる。また、文章の段落の性質を反映させることで、より正確に要約を作成する特開平3-191475号公報に記載された方法等も提案されている。

【0003】

【発明が解決しようとする課題】しかし、同一の文書でも、例えば営業用や技術資料用等の利用目的その他のユーザーの嗜好が異なると、文書における重要部位等に差異が生じる。そして、上述のような従来の文書処理によって要約を作成しても、ユーザーの嗜好を踏まえた要約を得ることはできない問題点がある。

【0004】本発明は、上述のような課題を解決するためになされたもので、利用目的等のユーザーの嗜好を踏まえた要約自動作成結合文書処理を行うことのできる文書処理装置、文書処理プログラムを記憶した記憶媒体、及び文書処理方法を提供することを目的とする。

【0005】

【課題を解決するための手段】請求項1に記載の発明は、図4に示すように、要約を作成する対象となる要約対象文書を取得する対象文書取得手段101と、ユーザーの嗜好を反映した嗜好資料を取得する嗜好資料取得手段102と、上記対象文書取得手段101により取得した要約対象文書と上記嗜好資料取得手段102により取得した嗜好資料とを結合して結合文書を取得する文書結

台手段103と、上記文書結合手段103により取得した結合文書から重要語句を抽出する重要語句抽出手段104と、前記重要語句抽出手段104により取得された重要語句を用いて前記要約対象文書から重要文を選択する重要文選択手段105と、前記重要文選択手段105により選択された重要文により前記文書の要約を作成する嗜好要約作成手段106と、を具備する文書処理装置を提供することにより、前記目的を達成する。請求項2に記載の発明は、図4に示すように、請求項1に記載の文書処理装置において、前記嗜好資料取得手段102は、嗜好資料として、ユーザーの嗜好を反映した文書またはプロフィールを使用する文書処理装置を提供することにより、前記目的を達成する。請求項3に記載の発明は、図5に示すように、請求項1または請求項2に記載の文書処理装置において、要約対象文書全体に対する要約の比率を取得する要約比率取得手段107を備え、前記重要文選択手段105は、前記要約比率取得手段107で取得した前記比率に従って重要文を選択する文書処理装置を提供することにより前記目的を達成する。請求項4に記載の発明は、図6に示すように、要約の作成対象となる要約対象文書を取得する対象文書取得機能201と、ユーザーの嗜好を反映した嗜好資料を取得する嗜好資料取得機能202と、上記対象文書取得機能201により取得した要約対象文書と上記嗜好資料取得機能202により取得した嗜好資料とを結合して結合文書を取得する文書結合機能203と、上記文書結合機能203により取得した結合文書から重要語句を取得する重要語句取得機能204と、前記重要語句取得機能204により取得された重要語句を用いて前記要約対象文書から重要文を選択する重要文選択機能205と、前記重要文選択機能205により選択された重要文により前記要約対象文書の要約を作成する嗜好要約作成機能206と、をコンピュータに実現させるためのコンピュータ読みとり可能な文書処理プログラムが記憶された記憶媒体を提供することにより、前記目的を達成する。請求項5に記載の発明は、図6に示すように、請求項4に記載の記憶媒体において、前記嗜好資料取得機能202は、嗜好資料として、ユーザーの嗜好を反映した文書またはプロフィールを使用する文書処理プログラムが記憶された記憶媒体を提供することにより前記目的を達成する。請求項6に記載の発明は、図7に示すように、請求項4または請求項5に記載の記憶媒体において、要約対象文書全体に対する要約の比率を取得する要約比率取得機能207を備え、前記重要文選択機能205は、前記要約比率取得機能207で取得した前記比率に従って重要文を選択する文書処理プログラムが記憶された記憶媒体を提供することにより前記目的を達成する。請求項7に記載の発明は、図8に示すように、要約の作成対象となる要約対象文書及びユーザーの嗜好を反映した嗜好資料を取得301し、上記要約対象文書と上記嗜好資料とを結合して

結合文書を取得302し、上記結合文書から重要語句を取得303し、前記重要語句を用いて前記要約対象文書から重要文を選択304し、この重要文により前記文書の要約を作成する305書処理方法を提供することにより前記目的を達成する。

【0006】

【発明の実施の形態】以下、本発明の文書処理装置、文書処理プログラムを記憶した記憶媒体、及び文書処理方法の好適な実施の形態について、図1から図3を参照して詳細に説明する。

(1) 実施形態の概要

本実施形態では、要約対象文書に、ユーザーの嗜好を反映した文書を結合し、得られた結合文書全体から形態素解析等により重要語の候補語を抽出する。そして、結合文書中における出現頻度等から、候補語 x の語句重要度 $f(x)$ を取得し、語句重要度の高い候補語を重要語とする。得られた重要語 a, b, \dots とその語句重要度 $f(a), f(b), \dots$ は要約対象文書のみから重要語や語句重要度を取得する場合よりも、ユーザーの嗜好の反映されたものとなる。そしてこの重要語 a, b, \dots 及び語句重要度 $f(a), f(b), \dots$ に基づいて、要約対象文書の各文の文重要度 $F(X)$ を取得し、文重要度 $F(X)$ の高い文をリストアップし、要約対象文書中の出現順に並べて、要約とする。

【0007】(2) 実施形態の詳細

図1は、本発明の文書処理装置の一実施形態であり、本発明の文書処理プログラムを記憶した記憶媒体の一実施形態の該プログラムが読み取られたコンピュータの構成を表したブロック図である。この図1に示すように、文書処理装置(コンピュータ)は、装置全体を制御するための制御部11を備えている。この制御部11には、データベース等のバスライン21を介して、入力装置としてのキーボード12やマウス13、表示装置14、印刷装置15、記憶装置16、記憶媒体駆動装置17、通信制御装置18、および、入出力I/F19、および、文字認識装置20が接続されている。制御部11は、CPU111、ROM112、RAM113を備えている。ROM112は、CPU111が各種制御や演算を行うための各種プログラムやデータが予め格納されたリードオンリーメモリである。

【0008】RAM113は、CPU111にワーキングメモリとして使用されるランダムアクセスメモリである。このRAM113には、本実施形態による要約作成処理を行うためのエリアとして、対象文書格納エリア1131、要約パラメータ格納エリア1132、重要語・重要度格納エリア1133、結合文書格納エリア1134、要約格納エリア1135、その他の各種エリアが確保されるようになっている。

【0009】対象文書格納エリア1131には、要約作成の対象となる文書(要約対象文書)が格納される。ま

たこの対象文書格納エリア 1131 には、本実施形態により取得された文重要度 F (X) が、要約対象文書の各文に対応させて格納される。要約パラメータ格納エリア 1132 には、操作者からの入力等により取得された要約パラメータの値または後述のデータ格納部 163 から読み込んだ要約パラメータのデフォルト値が格納される。操作者が入力する要約パラメータとしては、例えば、全文書に対する要約の比率 (1~99)、数量優先のある／なし、長単文の優先のある／なし、です／ます／であるの選択をする／しない、等の値が格納される。重要語・重要度格納エリア 1133 には、それぞれ、本実施形態により取得された重要語 (句も含む) 及びそれらの語句重要度が、互いに対応付けられて格納される。結合文書格納エリア 1134 には、本実施形態により要約対象文書とユーザーの嗜好を反映した資料 (嗜好資料) とを結合した結合文書が格納される。前記嗜好資料は、ユーザーの要約文書の利用目的等の記載された文章や、ユーザーのプロファイル等が用いられる。要約格納エリア 1135 には、本実施形態により取得された重要文が、要約作成対象文書における順番で格納される。

【0010】キーボード 12 は、かな文字を入力するためのかなキーやテンキー、各種機能を実行するための機能キー、カーソルキー、等の各種キーが配置されている。操作者が要約比率を入力する場合には、該要約比率はこのキーボード 12 から入力され、要約パラメータ格納エリア 1132 に格納される。マウス 13 は、ポインティングデバイスであり、表示装置 14 に表示されたキーやアイコン等を左クリックすることで対応する機能の指定を行う入力装置である。表示装置 14 は、例えば CRT や液晶ディスプレイ等が使用される。この表示装置 14 には、嗜好要約作成の対象となる文書の内容や、本実施形態により作成された嗜好要約等が表示されるようになっている。印刷装置 15 は、表示装置 14 に表示された文章や、記憶装置 16 の文書データベース 165 に格納された文書等の印刷を行うためのものである。この印刷装置としては、レーザプリンタ、ドットプリンタ、インクジェットプリンタ、ページプリンタ、感熱式プリンタ、熱転写式プリンタ、等の各種印刷装置が使用される。

【0011】記憶装置 16 は、読み書き可能な記憶媒体と、その記憶媒体に対してプログラムやデータ等の各種情報を読み書きするための駆動装置で構成されている。この記憶装置 16 に使用される記憶媒体としては、主としてハードディスクが使用されるが、後述の記憶媒体駆動装置 17 で使用される各種記憶媒体のうちの読み書き可能な記憶媒体を使用するようにしてもよい。記憶装置 16 は、仮名漢字変換辞書 161、プログラム格納部 162、データ格納部 163、文書データベース 165、図示しないその他の格納部 (例えば、この記憶装置 16 内に格納されているプログラムやデータ等をバックアッ

プするための格納部) 等を有している。プログラム格納部 162 には、本実施形態における嗜好要約作成処理プログラム等の各種プログラムの他、仮名漢字変換辞書 161 を使用して入力された仮名文字列を漢字混り文に変換する仮名漢字変換プログラム等の各種プログラムが格納されている。

【0012】データ格納部 163 には、要約パラメータのデフォルト値等の各種データが格納されている。要約パラメータのデフォルト値としては、例えば、全文書に対する要約の比率 = 「25%」や、日付時刻、価格情報、物理量 (サイズ、重量、温度等) 等の数量重視 = 「しない」や、URL (Uniform Resource Locator) 重視 = 「しない」、長単文の重視 = 「しない」や、です／ます／であるの選択 = 「しない」、等の値が格納されている。

【0013】文書データベース 165 には、仮名漢字変換プログラムにより作成された文書や、他の装置で作成されて記憶媒体駆動装置 17 や通信制御装置 18 から読み込まれた文書が格納される。この文書データベース 165 に格納される各文書の形式は特に限定されるものではなく、テキスト形式の文書、HTML (Hyper Text Markup Language) 形式の文書、JIS 形式の文書等の各種形式の文書の格納が可能である。

【0014】記憶媒体駆動装置 17 は、CPU 111 が外部の記憶媒体からコンピュータプログラムや文書を含むデータ等を読み込むための駆動装置である。記憶媒体に記憶されているコンピュータプログラムには、本実施形態の文書処理装置により実行される各種処理のためのプログラム、および、そこで使用される辞書、データ等も含まれる。ここで、記憶媒体とは、コンピュータプログラムやデータ等が記憶される記憶媒体をいい、具体的には、フロッピーディスク、ハードディスク、磁気テープ等の磁気記憶媒体、メモリチップや IC カード等の半導体記憶媒体、CD-ROM や MO、PD (相変化書換型光ディスク) 等の光学的に情報が読み取られる記憶媒体、紙カードや紙テープ等の用紙 (および、用紙に相当する機能を持った媒体) を用いた記憶媒体、その他各種方法でコンピュータプログラム等が記憶される記憶媒体が含まれる。本実施形態の文書処理装置において使用される記憶媒体としては、主として、CD-ROM やフロッピーディスクが使用される。記憶媒体駆動装置 17 は、これらの各種記憶媒体からコンピュータプログラムを読み込む他に、フロッピーディスクのような書き込み可能な記憶媒体に対して RAM 113 や記憶装置 16 に格納されているデータ等を書き込むことが可能である。

【0015】本実施形態の文書処理装置では、制御部 11 の CPU 111 が、記憶媒体駆動装置 17 にセットされた外部の記憶媒体からコンピュータプログラムを読み込んで、記憶装置 16 の各部に格納 (インストール) する。そして、本実施形態による類似度算出等の各種処理

を実行する場合、記憶装置16から該当プログラムをRAM113に読み込み、実行するようになっている。但し、記憶装置16からではなく、記憶媒体駆動装置17により外部の記憶媒体から直接RAM113に読み込んで実行することも可能である。また、文書処理装置によっては、本実施形態の嗜好要約作成処理プログラム等を予めROM112に記憶しておき、これをCPU111が実行するようにしてもよい。

【0016】通信制御装置18は、他のパーソナルコンピュータやワードプロセッサ等との間でテキスト形式やHTML形式等の各種形式の文書やビットマップデータ等の各種データの送受信を行うことができるようになっている。入出力1/F19は、音声や音楽等の出力を行うスピーカ等の各種機器を接続するためのインターフェースである。文字認識装置20は、用紙等に記載された文字をテキスト形式やHTML等の各種形式で認識する装置であり、イメージスキャナや文字認識プログラム等で構成されている。

【0017】本実施形態では、キーボード12の入力操作により作成した文書（RAM113の所定格納エリアに格納）の他、外部で作成して所定の記憶媒体に格納した文書で記憶媒体駆動装置17から読み込んだ文書、予め文書データベースに格納されている文書、通信制御装置18からダウンロードした文書、及び文字認識装置20で文字認識した文書、等の各種文書を対象文書として取得することが可能である。

【0018】次に、上述のような構成の文書処理装置による嗜好要約作成処理であって、本発明の文書処理方法の一実施形態について図2及び図3を参照して説明する。

【0019】図2は、本実施形態による嗜好要約作成処理のメイン動作を表すフローチャートである。嗜好要約作成処理に際しては、CPU111は、要約対象文書を取得し、RAM113の対象文書格納エリア1131に格納する（ステップ11）。要約対象文書は、ユーザの指示に従ってRAM113、記憶装置16の文書データベース165、記憶媒体駆動装置17、または通信制御装置18から取得する。また、CPU111は、上記要約対象文書と同様の手法により嗜好資料を取得し（ステップ12）、前記要約対象文書と前記嗜好資料とを結合させてRAM113の結合文書格納エリア1134に格納する（ステップ13）。次に、CPU111は、ユーザによってキーボード12等から要約パラメータが入力された場合には入力値を取得し、ユーザによる入力がない場合にはデータ格納部163に格納された要約パラメータのデフォルト値を取得し、要約パラメータ格納エリア1132に格納する（ステップ14）。続いて、CPU111は、結合文書についての重要語及びそれらの語句重要度を取得する（ステップ15）。

【0020】図3は、本実施形態における重要語・語句

重要度取得処理の動作を表したフローチャートである。図3に示すように、CPU111は、結合文書について、形態素解析を行うことで結合文書から自立語を抽出する（ステップ151）と共に、名詞句、複合名詞句等を含めた候補語（句）を結合文書から抽出する（ステップ152）。次に、RAM16の要約パラメータ格納エリア1132に格納した要約パラメータや、抽出した候補語（句）の結合文書での出現頻度、評価関数から、各候補語（句） x の語句重要度 $f(x)$ を取得する（ステップ153）。ここで、評価関数としては、例えば、所定の重要語が予め指定されている場合にはその重要語に対する重み付け、単語、名詞句、複合名詞句等の候補語（句）の種類による重み付け、等が使用される。

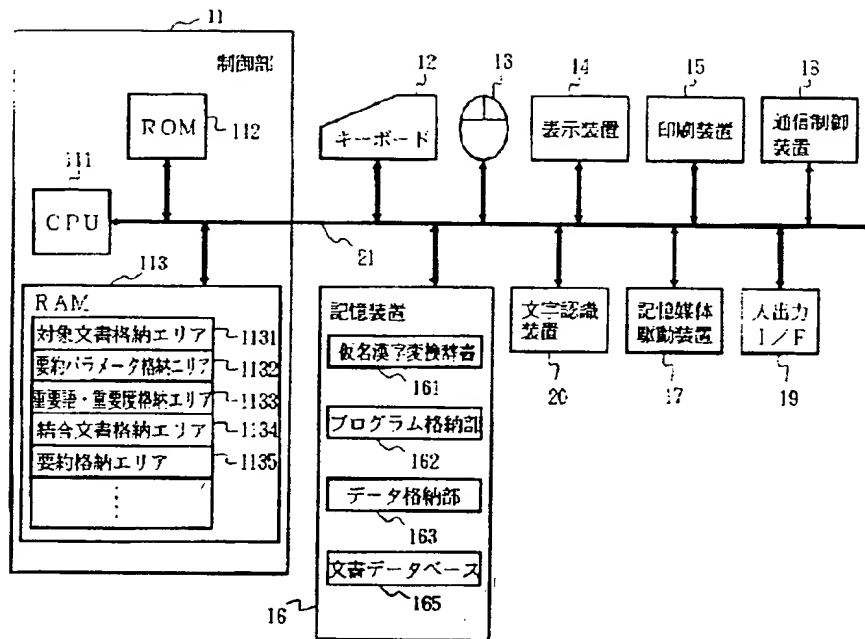
【0021】さらにCPU111は、取得した語句重要度 $f(x)$ の値をもとに候補語（句）から重要語 a 、 b 、 c 、…を取得し（ステップ154）、この重要語 a 、 b 、 c 、…及びその語句重要度 $f(a)$ 、 $f(b)$ 、 $f(c)$ …を重要語・重要度格納エリア1133に格納し（ステップ155）、図2に示す要約作成処理ルーチンへリターンする。

【0022】次に、CPU111は、重要語及びその語句重要度から、対象文書格納エリア1131に格納された要約対象文書の各文に対する文重要度 $F(X)$ を取得する（ステップ16）。この文重要度 $F(X)$ は、各文中における重要語の語句重要度を累積し、かつ文中において複合名詞句を検索し、複合名詞句による重み付けをして求める。そして、CPU111は、決定した各文の文重要度 $F(X)$ の高い文の上位から要約パラメータの要約比率（例えば、対象要約文書中の全文数の内の上位25%）以内に入る文（重要文）をリストアップし、要約格納エリア1137に格納する（ステップ17）。そして、リストアップした文を要約対象文書の中での出現順に並べることで当該要約対象文書の嗜好要約とし（ステップ18）、本実施形態による要約作成処理を終了する。

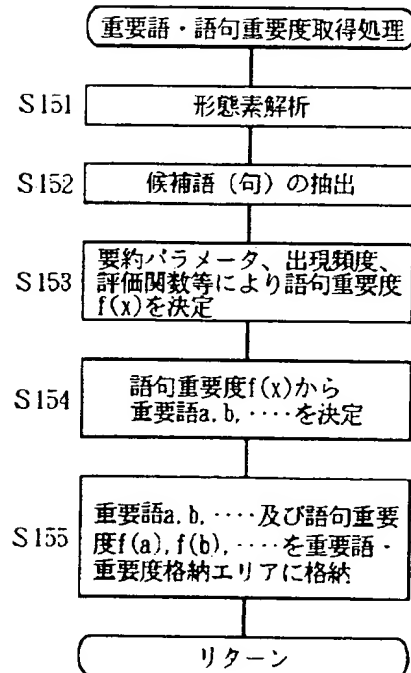
【0023】この様に、本実施形態では、要約対象文書にユーザの嗜好の反映された嗜好資料を結合し、得られた結合文書をもとに重要語 a 、 b 、…及び語句重要度 $f(a)$ 、 $f(b)$ 、…を取得し、この重要語 a 、 b 、…及び語句重要度 $f(a)$ 、 $f(b)$ 、…に基づいて要約対象文書中の各文の文重要度 $F(X)$ を取得し、重要文を決定する。従って、本実施形態によれば、ユーザの嗜好の反映された要約が作成される。また、本実施形態では、キーボード12からの入力により要約比率（要約対象文書全体に対する嗜好要約の比率）を1～99%で自由に設定でき、所望の分量の要約が作成できる。

【0024】尚、本発明は、上述の実施形態に限定されるものではなく、本発明の趣旨を逸脱しない限りにおいて適宜変更が可能である。例えば、上述の実施形態においては文書処理装置としてコンピュータを用いている

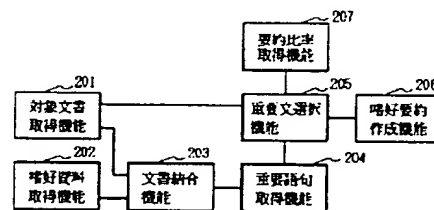
【図1】



【図3】



【図7】



【図2】

